Machine Learning Engineer Nanodegree

Capstone Proposal

Ahmed Al-Sabounchi – May 8th, 2019

Domain Background

Hand gesture recognition is becoming an attractive area of research as it provides the possibility of achieving touchless user interfaces which can be very useful in various industries as they enable the user to save time by eliminating the need to type a command or enter it using a touch interface, while maintaining a high level of accuracy. For example, it can be useful in the automotive industry as a means of improving safety and comfort. These touchless interfaces would allow the drivers to carry out commands while focusing on driving [1].

In order to achieve this task, Convolutional Neural Networks (CNNs) have been employed, where thousands of images of pre-labeled hand gestures are fed into the neural network to train it to recognize different hand gestures.

Problem Statement

The aim of this project is to create a CNN that is able to identify one of 10 different hand gestures. By achieving this, these gestures can then be integrated into a user interface that would link them to specific outputs. This would help make a fast and accurate user interface that saves the user time and adds safety to some of their essential daily tasks, such as driving.

Datasets and Inputs

The "Hand Gesture Recognition Database", provided by Leap Motion, will be used to train the CNN. This dataset, which was obtained from Kaggle, contains around 20,000 images of 10 individuals (5 male, 5 female) making 10 distinct hand gestures [2]. The first image in this dataset are labelled as follows: frame_00_01_0001, where the 00 corresponds to the individual's identifier (0 to 9), the 01 corresponds to the hand gesture (1 to 10), and 0001 corresponds to the specific image of the gesture made by that individual (there are 200 images taken of the same gesture by each individual, to add for translational invariance).

For this project, the images will be grouped based on the gesture. This means that pictures of different individuals performing the same action will be grouped together. The model will read the two digits corresponding to the specific gesture then translate it into text.

Solution Statement

A CNN will be designed to perform this task, where different available pre-trained models in Keras will be tested to find the best one for this application, followed by a modification in the final layer, where a Global Average Pooling layer will be used to flatten the data, and a Dense layer will be used to classify the image into one of the 10 gesture categories.

Benchmark Model

The chosen benchmark model is a hand gesture recognition model that employs Gabor filters and Support Vector Machine (SVM) classifiers. This model consists of a segmentation stage that determines where the hand is placed in the image, then computes image descriptors using Gabor filters, before finally using SVM classification for gesture recognition [3].

The idea is that the model would yield the same output type as the proposed solution, making it appropriate for evaluation against it.

Evaluation Metric

A Confusion Matrix will be used to obtain the Precision and Recall values for this model. With precision defined as $Precision = 100 \times \frac{True\ Positives}{True\ Positives + False\ Positives}$, this will be the metric that will be used to quantify the accuracy of the proposed model and compare it with the benchmark model.

Project Design

The first step would be to construct the benchmark model and pass the given dataset through it to obtain the benchmark numbers. Then, the proposed CNN will be constructed, starting with Transfer Learning, where a pre-trained model will be used for the convolutional layers. The particular model will be chosen based on the accuracy obtained through trial and error. This model will be capped off with a global average pooling layer to flatten the data before passing it through a Dense layer with 10 nodes and a SoftMax activation function. Finally, the model will be evaluated against the benchmark model based on the proposed evaluation metric.

References

[1] P. Molchanov, S. Gupta, K. Kim and J. Kautz, "Hand gesture recognition with 3D convolutional neural networks", *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2015. Available: 10.1109/cvprw.2015.7301342 [Accessed 8 May 2019].

[2] "Hand Gesture Recognition Database", *Kaggle.com*, 2019. [Online]. Available: https://www.kaggle.com/gti-upm/leapgestrecog. [Accessed: 08- May- 2019].

[3] D. Huang, W. Hu and S. Chang, "Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination", *Expert Systems with Applications*, vol. 38, no. 5, pp. 6031-6042, 2011. Available: 10.1016/j.eswa.2010.11.016 [Accessed 9 May 2019].